JOURNAL OF VIROLOGY, Mar. 2006, p. 2349–2357 Vol. 80, No. 5
0022-538X/06/$08.00+0 doi:10.1128/JVI.80.5.2349–2357.2006

Distinct Viral Populations Differentiate and Evolve Independently in a Single Perennial Host Plant†

Chiraz Jridi,[1] Jean-François Martin,[2] Véronique Marie-Jeanne,[1] Gérard Labonne,[1] and Stéphane Blanc[1]*

UMR Biologie et Génétique des Interactions Plantes-Parasites, CIRAD-INRA-ENSAM, TA 41/K, Campus International de Baillarguet, 34398 Montpellier Cedex 05, France,[1] and UMR Biologie et de Gestion des Populations, CIRAD-ENSAM-INRA-IRD, Campus International de Baillarguet, 34398 Montpellier Cedex 05, France[2]

Received 27 September 2005/Accepted 6 December 2005

The complex structure of virus populations has been the object of intensive study in bacteria, animals, and plants for over a decade. While it is clear that tremendous genetic diversity is rapidly generated during viral replication, the distribution of this diversity within a single host remains an obscure area in this field of science. Among animal viruses, only *Human immunodeficiency virus* and *Hepatitis C virus* populations have recently been thoroughly investigated at an intrahost level, where they are structured as metapopulations, demonstrating that the host cannot be considered simply as a "bag" containing a homogeneous or unstructured swarm of mutant viral genomes. In plants, a few reports suggested a possible heterogeneous distribution of virus variants at different locations within the host but provided no clues as to how this heterogeneity is structured. Here, we report the most exhaustive study of the structure and evolution of a virus population ever reported at the intrahost level through the analysis of a *Prunus* tree infected by *Plum pox virus* for over 13 years following a single inoculation event and by using analysis of molecular variance at different hierarchical levels combined with nested clade analysis. We demonstrate that, following systemic invasion of the host, the virus population differentiates into several distinct populations that are isolated in different branches, where they evolve independently through contiguous range expansion while colonizing newly formed organs. Moreover, we present and discuss evidence that the tree harbors a huge "bank" of viral clones, each isolated in one of the myriad leaves.

Viruses are notable for their large census population size, short generation time, and high replication and mutation rates. In fact, because of the lack of proofreading activity associated with RNA-dependent RNA polymerase, RNA viruses have the highest mutation rates among living entities (8), with extremely high genetic variability being generated rapidly within viral populations (7). Analysis of the genetic structure and evolution of populations is a crucial area of general biology and, in the case of viruses, highly relevant to the development of strategies for disease or epidemic control. Over the last 20 years, interest in the genetic structure and evolution of virus populations has emerged and developed tremendously, as evidenced by the ever-increasing number of papers published on this subject (for recent reviews, see references 18, 21, 28, and 39).

The concept of the viral population has been hotly debated around the term quasispecies (6, 12, 22, 47), but it is undisputed that an RNA virus population is a swarm of mutant genomes among which complementation occurs to a variable extent for different vital functions (16). The limit or frontier of virus populations is difficult to determine and varies widely depending on the scientific questions addressed by various authors. This limit is often logically determined by the physical or geographical barriers that separate host populations. Infected pluricellular hosts have been, and are still, often considered as delineating the minimal virus population since no clear isolation is perceived (or known) among groups of viral genomes replicating in various locations within this host, and mixing can occur through the vascular system, resembling a panmictic situation. Consequently, in animals and most particularly in plants, the structure of virus populations is seldom described in detail at an intrahost level. When sequence analysis of a genome pool originating from a single host is reported, this pool is usually considered a unique population sample and is rarely regarded as a mix of several possible bulk genomes originating from various organs and tissues (14, 19, 45).

The few data available in the literature at this intrahost level suggest a more complex pattern than the genuine panmictic situation. Remarkably, for *Human immunodeficiency virus* type 1 (HIV-1), it was shown that the assumption that populations are panmictic within a host is inconsistent with the observed data. Instead, the metapopulation concept, where a population is made up of several discrete subpopulations that undergo turnover through frequent foundation and extinction events, could play a central role in the evolution of HIV-1 (17). Differentiation of subpopulations in various organs of the same host has been further confirmed more recently for HIV (24, 30, 33, 37) and has also been reported for other animal virus species such as *Hepatitis C virus* (1, 9, 25) and *TT virus* (29).

In plants, available information is scarcer. Analysis of the genetic composition of virus populations during systemic invasion of tobacco plants by *Tobacco mosaic virus* (36) and *Cucumber mosaic virus* (26) or wheat plants by *Wheat streak mosaic virus* (15, 20) has clearly revealed the existence of severe

* Corresponding author. Mailing address: UMR Biologie et Génétique des Interactions Plantes-Parasites, CIRAD-INRA-ENSAM, TA 41/K, Campus International de Baillarguet, 34398 Montpellier Cedex 05, France. Phone: 33 (0)4 99 62 48 04. Fax: 33 (0)4 99 62 48 22. E-mail: blanc@ensam.inra.fr.

† Supplemental material for this article may be found at http://jvi.asm.org/.

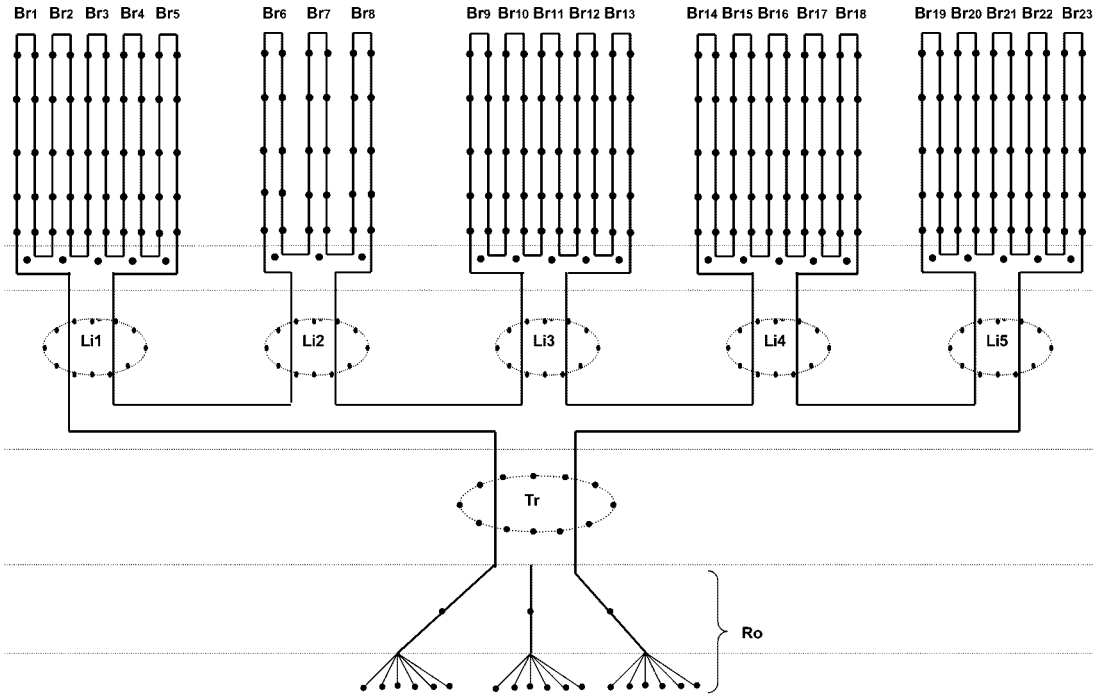

FIG. 1. Sample locations within the *Prunus persica* host. The architecture of the tree is drawn schematically, comprising terminal roots, main roots, trunk, constitutive limbs, yearling branches, and leaves. An extensive sampling was conducted according to the stratified pattern as follows: dots situated at the periphery of each of the upper branches indicate whole-leaf samples; the dot situated at the base of each of the upper branches indicates a branch bark sample; dotted circles illustrate virtual transverse sections, around which each dot represents one bark sample; and dots on the main and young terminal roots indicate bark and whole tissue samples, respectively. Groups of samples Br1 to Br23, Li1 to Li5, Tr, and Ro are hypothesized to contain distinct subpopulations (see text). Samples from the main and young terminal roots were grouped in a single Ro subpopulation.

bottlenecks when a virus population is colonizing new leaves, possibly inducing strong genetic drift and differentiation within each leaf, related to the founder effect. Perennial plants are particularly well suited as models for the study of the structure of virus populations. Indeed, infections commonly persist for many years without killing the host, and the viral population can potentially be analyzed after a much longer within-host evolution than in annual plants. It is thus surprising that very few studies have investigated the structure of virus populations within perennial hosts. Previous hints, however, indicate that the composition of sequence variants (also designated haplotypes) may differ in various locations of a single host tree for *Apple stem grooving virus* (27) and *Citrus tristeza virus* (10), although other data appear to be contradictory (11).

To further study the differentiation of several subpopulations of plant viruses at the intrahost level and to elucidate the mechanisms of such differentiation, we have decided to use a perennial plant model chronically infected by a potyvirus following a unique inoculation event. The extensive analysis presented in this report reveals a series of striking phenomena that have thus far been mostly overlooked in plant virology. Thirteen years after the initial inoculation, we demonstrate that a high genetic diversity has built up in the virus population, which is structured by the architecture of the tree. Beyond the observation that viral genetic diversity increases when moving up from old (trunk and then limb) to newly formed (branches and then leaves) organs, our data clearly demonstrate that several viral subpopulations differentiate over the years as they become isolated in different limbs of the host tree and evolve independently by "contiguous range expansion." In addition, we present evidence that each individual leaf is infected by a single haplotypic variant; thus, the tree harbors, through its myriad leaves, a huge collection of mutant genomes comprising the complex sum of a virus population(s).

MATERIALS AND METHODS

Virus and plant materials. The host-virus pair chosen in this study involves a single peach tree, *Prunus persica* cv. Alexandra, that has been infected with *Plum pox virus* (PPV) strain M since 1991. More detailed information on the organization of the RNA genome of PPV is available in the supplemental material (see Fig. S1 in the supplemental material). PPV strain M isolate 91-001, used as an inoculum, had previously been maintained in pea plants by sap inoculation transfer. In the spring of 1991, aphids, previously starved for 2 h, were placed on an infected pea leaf for a 7-min acquisition access period and immediately transferred onto a single leaf on the young *Prunus* tree (1 year old, single-shoot stage) for an overnight inoculation access period. Since then, the infected *Prunus* tree has been grown and kept in an insect-proof screened house to prevent any possibility of secondary infection.

Sampling protocol. In June 2004, 13 years after PPV inoculation, we carried out a large-scale stratified sampling on the infected *Prunus* tree. As schematized in Fig. 1, we collected the following samples on the tree: (i) 3 main roots (bark only) and 18 young terminal roots (pieces of well-separated fibrous root); (ii) 12 bark samples from the trunk, evenly distributed around the periphery of a virtual transverse section; (iii) 12 bark samples from the base of each of the five constitutive limbs, evenly distributed at the periphery of virtual transverse sections; (iv) 1 bark sample from the base of each of 23 newly formed branches, with a maximum of 5 branches (when available) being sampled on each of the five

constitutive limbs; and (v) 10 whole-leaf samples (when available) from each of the 23 young branches. In total, 230 whole leaves, 23 bark samples of young branches, 60 bark samples of limbs, 12 bark samples of the trunk, 3 bark samples of main roots, and 18 young terminal roots were collected. All samples were stored at −80°C until use for immunocapture reverse transcription (RT)-PCR and single strand conformation polymorphism (SSCP) analysis.

cDNA synthesis and PCR amplification. Three genomic regions, designated A, B, and C, corresponding to sequences encoding the N terminus of HC-Pro (nucleotide positions 1136 to 1371); a region overlapping the C terminus of P3, all of 6K1, and the N terminus of CI (positions 3425 to 3728); and a region overlapping the C terminus of Nib and the N terminus of CP (positions 8466 to 8891), respectively, were analyzed. The genome organization of PPV, summarizing the known functions of its different genes and the positioning of regions A, B, and C, is given in the supplemental material (see Fig. S1 in the supplemental material). Three pairs of primers (A, B, and C) were designed for performing RT-PCR for these three regions of the PPV genome; primer sequences are also available in supplemental material (see Table S2 in the supplemental material).

Plant extracts were prepared by grinding leaves, barks, and fragments of roots 1/20 (wt/vol) in phosphate-buffered saline (PBS)–Tween (PBS containing 0.05% [vol/vol] Tween 20 and 2% [wt/vol] polyvinylpyrrolidone). For each sample, 100 μl of supernatant from a low-speed centrifugation (12,000 × *g*, 5 min) was incubated overnight at 4°C in tubes previously coated with PBS-Tween containing 1 μg ml^{-1} polyclonal anti-PPV immunoglobulin G; the supernatant was then washed away with two rinses with the same buffer. RT was performed directly in these immunocapture tubes using a reverse transcription kit (Promega, Madison, WI), according to the supplier's recommendations. One microliter of the cDNA-containing mixture was then used for further PCR amplification. The yields of the resulting RT-PCR products were systematically checked on agarose gels and then stored at a concentration of 100 ng μl^{-1}.

SSCP analysis and sequencing. For SSCP analysis, 5 μl (0.5 μg) of each of the RT-PCR products was mixed with 12 μl of denaturing solution (95% formamide, 20 mM EDTA, pH 8.0, 0.05% bromophenol blue, and 0.05% xylene cyanol), heated for 10 min at 95°C, and immediately cooled on ice. The DNA strands were separated by nondenaturing 8% polyacrylamide gel electrophoresis in Tris-borate-EDTA buffer (89 mM Tris-borate, pH 8.0, 2 mM EDTA), with a constant current of 28 mA per minigel, and the temperature was maintained at 4°C. The duration of the electrophoresis run varied with the size of the fragment analyzed and was 4, 5, and 6 h for regions A, B, and C, respectively. The gels were finally stained with a DNA silver staining kit (Amersham Biosciences, Little Chalfont, United Kingdom).

Tissue samples yielding different SSCP patterns in at least one of the genomic regions analyzed were considered to be containing different haplotypes. On the one hand, sequencing of all different haplotypes confirmed that they indeed differ by one or more mutations; on the other hand, sequencing of up to five samples (when available) containing identical haplotypes confirmed that they displayed strictly identical sequences.

Fragmentation of the data set for statistical analysis. Samples from each of the 23 newly formed branches (10 leaf samples and one bark sample per branch, named Br1 to Br23), from each of the five constitutive limbs (12 bark samples per limb, named Li1 to Li5), from the trunk (12 bark samples, named Tr), and from the whole root system (named Ro) were a priori considered to be containing discrete population units. Thus, in total, we primarily hypothesized that we were working with 30 putatively distinct subpopulations positioned as shown in Fig. 1.

Descriptive statistics of genetic structure. Genetic differentiation between the 30 PPV subpopulations defined above was estimated by the *F* statistic (46), using the Arlequin version 2.000 program, and the significance was tested by bootstrap analysis based on 10,000 replicates.

The potential effect of a limb that supports branches and leaves on the total variation of haplotype distribution was studied by analysis of molecular variance (AMOVA) (13). This procedure was performed by evaluating the type and frequency of haplotypes detected in each limb (including associated branches and leaves) and by estimating the probability that their distribution was random. This test was contrasted using a bootstrap analysis with 10,000 repetitions. The statistic for differentiation between groups of subpopulations on different limbs was F_{CT}, the significance of which is similar to the corresponding percentage of variance. The second level of variation was between subpopulations within each limb. This source of variation was tested by exchanging the haplotypes among Br and Li subpopulations in each limb, and the statistic of differentiation was F_{SC}, the significance of which is also evaluated as a percentage of the corresponding variance. The third level of variation was within subpopulations, and the statistic of differentiation was F_{ST}, the significance of which is also evaluated as a percentage of the corresponding variance.

For neutrality testing, the program DnaSP (34) was used to study Tajima's *D* and Fu and Li's F^* statistics.

Nested clade analysis. We used a nested clade analysis (NCA) to evaluate the association between haplotypes and their geographical range. The NCA was performed according to the procedure described previously by Templeton et al. (43) and can be divided into a series of successive steps implemented with the computer programs TCS version 1.13 (2) and Geodis (32). First, a haplotype tree or network was constructed according to a statistical parsimony procedure (23, 40). The same TCS program was used to reconstruct a nonrooted haplotype genealogy and to identify the most ancestral haplotype. Second, the nesting procedure (3, 4, 40, 42, 44) was applied to the haplotype network. The resulting nested clades are designated by C-N, where C is the nesting level of the clade and N is the number assigned to a particular clade at a given nesting level (41). Once the cladogram had been converted into a nested series of linked clades, the geographical data were quantified via a distance matrix (in centimeters) based on the tree (host) structure. Finally, the computation of different distance statistics (D_c, the geographical range of a particular clade, and D_n, the distribution of a particular clade relative to its sister clades) was performed using Geodis (32) according to the procedures described previously by Templeton et al. (43).

In cases where clades were nonrandomly distributed, the biological causes of the haplotype-geography association were interpreted using the inference key previously proposed by Templeton et al. in 1995 (43) and updated recently (31). This inference key aims to discriminate between distinct evolution patterns throughout the hierarchical clade levels. These patterns are designated as restricted gene flow, past fragmentation, and range expansion.

RESULTS

To study the genetic structure of a plant virus population at the intrahost level, we extensively sampled a *Prunus* tree infected with PPV strain M following a single inoculation event that occurred more than 13 years ago. The sampling was carefully designed to show the possible relationships between the distribution of virus variants, the nature of the tree organs, and/or the architecture of the tree (Fig. 1) (see Materials and Methods for details). Three genomic regions (A, B, and C) of the PPV genome were analyzed by SSCP and DNA sequencing (sequences are available upon request), and each identified haplotype was characterized by a combination of these three regions. Thirty-three distinct haplotypes were identified, indicating a large within-host variation for PPV. Among the 333 samples analyzed, the frequency of the 33 haplotypes was highly variable. Two haplotypes, H1 and H4, were by far the most frequent, being found in 39.9 and 30.9% of the samples, respectively. Thirteen additional haplotypes were found in more than one sample, ranging from 2 to 20 samples, whereas 18 haplotypes were detected only once. Table 1 lists the 33 haplotypes, their frequencies, and the origin of the samples in which they were found. At this point, it should be mentioned that no selection for or against any of the 33 haplotypes was detected. Indeed, Tajima's *D* and Fu and Li's *F* statistics (see Materials and Methods) gave nonsignificant negative values not only on the complete data set (−1.14749 and −0.94357, respectively; $P < 0.1$) but also on the five groups of samples compared in the AMOVA analysis (data not shown) described below. Consequently, any heterogeneity of the haplotype distribution within this tree could not be attributed to local adaptation (selection) but likely was due to stochastic phenomena such as genetic drift and/or founder effect.

In order to intuitively evaluate whether and how the viral genetic diversity is structured within the host tree, and to elaborate working hypotheses, we represented each haplotype as a thin rectangle composed of three distinct regions of different colors, illustrating variations in regions A, B, and C, and

TABLE 1. Distribution of the 33 haplotypes within samples composing the 23 putative subpopulations

Haplotype	No. of samples containing haplotype:[a]																														
	Ro (21)	Tr (12)	Li1 (10)	Br1 (11)	Br2 (10)	Br3 (11)	Br4 (11)	Br5 (11)	Li2 (11)	Br6 (11)	Br7 (11)	Br8 (11)	Li3 (10)	Br9 (11)	Br10 (11)	Br11 (11)	Br12 (11)	Br13 (11)	Li4 (12)	Br14 (11)	Br15 (11)	Br16 (11)	Br17 (10)	Br18 (10)	Li5 (11)	Br19 (11)	Br20 (11)	Br21 (11)	Br22 (7)	Br23 (11)	Total (333)
H1	18	12							10	5	8	5	9		10	5	2		12	5	7	8	5	10	1				1		133
H2	2																				1										3
H3	1										1					4				1											7
H4			10	9	8	9	10	10																	9	8	11	2	6	11	103
H5				1																											1
H6				1																											1
H7					1																										1
H8					1	1	1	1																							4
H9						1																									1
H10									1	3																					4
H11										2																					2
H12										1										1											2
H13											1		1				6	11				1									20
H14											1																				1
H15												6																			6
H16														9		2															11
H17														1																	1
H18														1																	1
H19															1																1
H20																	1														1
H21																	1														1
H22																	1														1
H23																				2	1	1									4
H24																				1	1										2
H25																				1											1
H26																					1										1
H27																						1									1
H28																							2								2
H29																							1								1
H30																							1								1
H31																							1								1
H32																									1	2		9			12
H33																										1					1

[a] Numbers in the table represent the numbers of samples containing a given haplotype. Name of putative subpopulations are defined in the legend of Fig. 1. Numbers in parentheses are total numbers of samples successfully analyzed in each subpopulation.

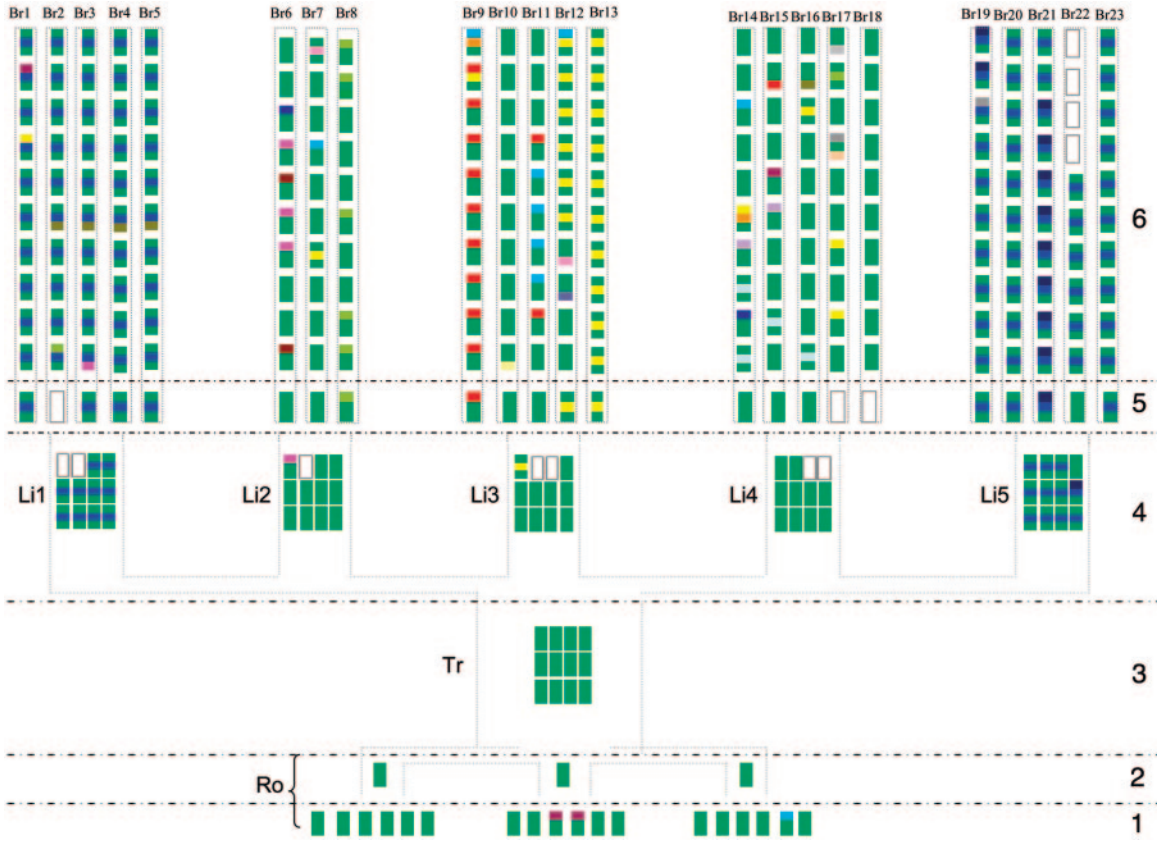

FIG. 2. Haplotype distribution superimposed on the schematic architecture of the tree. Each haplotype is represented as a thin rectangle composed of three distinct compartments illustrating the different sequences of regions A, B, and C. To facilitate the schematic representation of haplotypes, the first detected haplotype, H1, for regions A, B, and C is represented in green. When the sequence of a given genomic region varies, the color of the corresponding compartment changes. Blank haplotypes correspond to samples from which the immunocapture RT-PCR failed repeatedly. (1) Haplotypes were detected in the 18 young terminal root samples. (2) Haplotypes were detected in the three main root samples. (3) Haplotypes were detected in the 12 bark samples from the trunk. (4) Haplotypes were detected in the 60 bark samples from constitutive limbs. (5) Haplotypes were detected in the 23 bark samples from yearling branches. (6) Haplotypes were detected in the 230 leaf samples.

superimposed the haplotypic distribution on the architecture of the host tree (Fig. 2). This representation immediately suggests a number of interesting hypotheses that can all be tested using the series of statistical tools described in Materials and Methods: (i) the haplotypic diversity is distributed heterogeneously and increases when sampling is done towards the newly formed organs of the tree, going from the trunk to the major limbs and then to branches and leaves; (ii) even in comparable organs, the high genetic diversity of the PPV population appears to be structured according to the tree architecture, with each supporting limb containing seemingly differentiated subpopulations; (iii) the distribution of apparent lines of PPV haplotypes seems to match the architecture of the tree, suggesting that the virus population has expanded in a "one-way" fashion, concomitant with tree growth and the appearance of new organs and tissues, with potentially no or little return and mixing. A detailed description and thorough statistical testing of these hypotheses are presented below.

Distribution of haplotype variants is highly heterogeneous within the host tree. The first striking observation is that in all samples analyzed and for all three genomic regions (A, B, and C), the SSCP profiles clearly demonstrated the presence of a single detectable sequence variant (data not shown). Despite the large genetic diversity found in our experimental tree, we never detected a mixture of haplotypes in any given sample. This surprising finding has very important implications that will be further discussed below.

To test for possible differentiation of PPV populations within this single host tree, the whole sample set was divided into 30 putative subpopulations according to the architecture of the tree, with 23 representing subpopulations from each yearling branch (Br1 to Br23), 5 representing subpopulations from the bark of each major limb (Li1 to Li5), and 2 representing subpopulations from the bark of the trunk and from the root system (Tr and Ro, respectively). We estimated the haplotype (H) and nucleotide (π) diversity within each of the 30 putative subpopulations (Table 2). The values were highly variable and ranged from 0.0 to 0.800 ($\pm$0.114) and from 0.0 to 0.0240 ($\pm$0.00054) for H and π, respectively. By comparing H and π obtained from bulk samples from the trunk, the limbs, the branches, and the leaves (Table 2), it appears that the diversity increases significantly when sampling is done upward in the host tree from the trunk to the leaves, from older to younger organs. Our sampling of the root system was only exploratory and not as extensive as that of the rest of the tree. For this reason, the variability found in young fibrous roots is

TABLE 2. Haplotype and nucleotide diversity within PPV putative subpopulations

Population[a]	Haplotype diversity (H)[b]	Nucleotide diversity (π)[b]
Ro	0.267 ± 0.120	0.00025 ± 0.00012
Tr	0	0
Li 1	0	0
Br1	0.345 ± 0.172	0.00033 ± 0.00018
Br2	0.378 ± 0.181	0.00092 ± 0.00048
Br3	0.345 ± 0.172	0.00067 ± 0.00040
Br4	0.182 ± 0.144	0.00050 ± 0.00039
Br5	0.182 ± 0.144	0.00050 ± 0.00039
Li2	0.182 ± 0.144	0.00050 ± 0.00039
Br6	0.745 ± 0.098	0.00200 ± 0.00047
Br7	0.533 ± 0.180	0.00102 ± 0.00040
Br8	0.545 ± 0.072	0.00100 ± 0.00013
Li3	0.200 ± 0.154	0.00037 ± 0.00028
Br9	0.345 ± 0.172	0.00110 ± 0.00062
Br10	0.182 ± 0.144	0.00017 ± 0.00013
Br11	0.691 ± 0.086	0.00077 ± 0.00015
Br12	0.709 ± 0.137	0.00160 ± 0.00040
Br13	0	0
Li4	0	0
Br14	0.800 ± 0.114	0.00240 ± 0.00054
Br15	0.618 ± 0.164	0.00130 ± 0.00047
Br16	0.491 ± 0.175	0.00127 ± 0.00051
Br17	0.756 ± 0.130	0.00185 ± 0.00061
Br18	0	0
Li5	0.318 ± 0.164	0.00076 ± 0.00043
Br19	0.473 ± 0.162	0.00093 ± 0.00036
Br20	0	0
Br21	0.327 ± 0.153	0.00060 ± 0.00028
Br22	0.286 ± 0.196	0.00079 ± 0.00054
Br23	0	0
Total tree	0.739 ± 0.017	0.00203 ± 0.00007
Total leaves	0.797 ± 0.019	0.00226 ± 0.00010
Total branch bark	0.749 ± 0.068	0.00191 ± 0.00025
Total limb bark	0.544 ± 0.038	0.00152 ± 0.00013
Total trunk bark	0	0

[a] Names of putative subpopulations.
[b] H, π, and standard variations were calculated using the program DnaSP (34).

not further considered in the detailed analysis described below and will be discussed later.

PPV subpopulations differentiate in various locations of the host tree. Genetic differentiation between the 30 putative subpopulations was assessed using the *F* statistic. In this test, out of 435 possible pairwise comparisons, 301 showed a significant difference (see Table S1 in the supplemental material), thus confirming that the haplotypic distribution is not uniform, even among comparable organs such as the 23 yearling branches (Br1 to Br23). Indeed, out of the 301 comparisons showing a significant differentiation, 179 pairwise comparisons were between different yearling branches.

In order to evaluate whether the differentiation detected between the subpopulation units is structured, and how it is structured, we performed an AMOVA test on the haplotype distribution at different hierarchical levels: the major limbs, the yearling branches within each limb, and the leaves within each yearling branch. While the AMOVA results (Table 3) showed significant genetic structure at the three hierarchical levels examined, the limb hierarchical level appears to explain the majority of the variation (55.04%). Differences between yearling branches of the same limb contributed to only 13.05% of the total variation. The remarkably large residual variation of 31.9% of the total is due to the high polymorphism within the 23 yearling branches (discussed in detail below).

PPV populations evolve by contiguous range expansion in the infected host tree. To further understand the structure and the evolutionary process of the PPV population within its host tree, we conducted an NCA on our data set (see Materials and Methods). When the haplotype phylogeny network was created, the Hudson estimator (23, 40) did not reject the hypothesis of parsimony for the entire data set, meaning that all haplotype sequences are tightly related and can be included in the same network (Fig. 3).

The nesting procedure also illustrated in Fig. 3 organized the haplotype network into a total of 15 step 1 level clades, nested in 6 step 2 level clades, which are themselves nested in 2 step 3 level clades. While the high number (15) of step 1 level clades further indicates a strong genetic differentiation within the PPV population, the fact that only two nesting levels can satisfactorily organize the network agrees with the close relatedness of all haplotype sequences.

Clade 1-1 contains the majority of the samples (141/333) and, in particular, samples containing the haplotype H1 (133/333), which is proposed to be the root of the network.

We measured the geographical distance (in centimeters) between the positions of all samples in the *Prunus* tree and confronted the haplotype network with the distance matrix. A significant nonrandom geographical association was found for 10 clades out of 20 clades tested. According to Templeton's inference key (31, 43), the inferred evolutionary process (Table 4), which explains the geographical association of clades, was contiguous range expansion for clades 1-1, 1-4, 1-6, and higher clade 2-2 and allopatric fragmentation for clade 1-7. However, for reasons discussed below, we believe the case of clade 1-7 is

TABLE 3. AMOVA within the PPV sample set

Hierarchical level	df[d]	Sum of squares	Components of variance	% Variation[e]	*F* statistic
Variation between limbs[a]	4	178.613	0.71027	55.04*	$F_{Ct} = 0.29037$
Variation between subpopulations within limbs[b]	23	50.912	0.16844	13.05*	$F_{SC} = 0.68098$
Variation within subpopulations[c]	272	111.971	0.41166	31.90*	$F_{ST} = 0.55044$
Total	299	341.497	1.29037		

[a] All bark and leaf samples supported by one of the five constitutive limbs are grouped.
[b] All Br and Li subpopulations were considered.
[c] All leaf and bark samples were considered, except those of Tr and Ro subpopulations.
[d] df, degree of freedom.
[e] *Significant ($P < 0.05$).

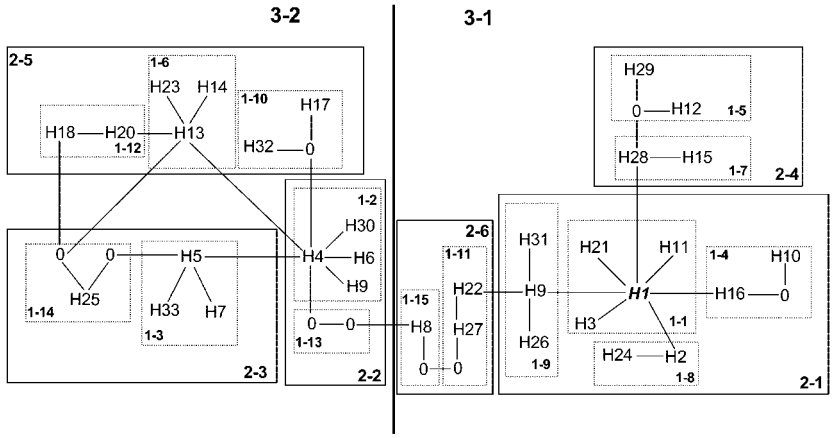

FIG. 3. Maximum parsimony network and nested clades of the 33 PPV haplotypes. Haplotypes are named as described in Table 1, and each connecting line represents a single mutational step between two haplotypes. Missing (hypothetical) haplotypes are represented by a "0." Dotted-line rectangles enclose step 1 level clades and are designated by "1-*n*"; plain-line rectangles enclose step 2 level clades and are designated "2-*n*"; and the thick line separates step 3 level clades (3-1 and 3-2). Haplotype H1 has the highest outgroup probability, as indicated by the TCS program.

to be considered with caution. Finally, no conclusive inference could be determined for clade 1-2 and higher clades 2-1, 2-5, 3-1, and 3-2.

DISCUSSION

Our biological system, a *Prunus persica* tree infected with PPV for 13 years, is a unique case of a controlled long-term plant/virus association. The fact that this tree has been protected from any possible secondary virus entry since the initial inoculation event makes it an excellent model for the basic study of the structure and evolution of a virus population within a single host. The use of SSCP as the primary technique to detect genetic variation in various locations of the host allowed the analysis of a large number of samples, thus providing the most thorough analysis of the genetic structure of a virus population ever reported at the intrahost level.

The observation that a unique sequence variant is detected in each of the very numerous samples is remarkably novel and has considerable implications. For main roots, trunk, and constitutive limbs, both the number of bark samples and the haplotypic diversity are low, and mixes could easily be overlooked under such conditions. In contrast, in leaves, the genetic diversity is very large, and up to 230 samples were analyzed. This situation was also confirmed in unrelated experiments, where single-sequence PPV variants were found in each of 450 additional whole-leaf samples from experimentally inoculated *Prunus* trees (C. Jridi, unpublished results). Previous reports have demonstrated the existence of strong genetic bottlenecking during the invasion of leaves by plant virus populations (15, 20, 26, 36). However, these studies suggested that more than one infectious unit was most often initiating the within-leaf population. To make sure that the surprising observation reported here was not an artifact of our experimental SSCP conditions, we pooled leaf samples by pairs and verified that haplotype mixes were readily detectable (data not shown). Moreover, we conducted SSCP on variable ratios of two PPV sequence variants in mixed solutions and observed that a sequence variant present at a frequency of 0.1 is still detectable (data not shown), a value also reported previously (35). From the latter test, we concluded that each leaf contains one major sequence with no other variant reaching a frequency of 0.1. Even in branches where two haplotypes are found with a high frequency among the 10 corresponding leaves (i.e., Br6, Br8, and Br11), no mix was ever detected within a single leaf. The only logical conclusion from these results is that the virus population is submitted to biological cloning at most (if not all) leaf invasions, and the tree thus harbors a tremendous number of clones within the complex virus population, each individually isolated in a single leaf. In apparent contradiction, a previous study on PPV has reported mixes of sequence variants within single leaves (5). In this study, healthy plants were coinoculated with two variants that were then colonizing "free-land" cells and tissues at their own pace, obviously reaching some leaves concomitantly. Interestingly, in an analysis of coinfection at the cellular level, those same authors reported that single cells were rarely (if ever) coinfected, thus suggesting that

TABLE 4. Evolution process of PPV populations within a single host tree

Clade[a]	Inference chain[b]	Inferred evolution process[c]
Haplotypes nested in 1-1	1-2-11-12-NO	Contiguous range expansion
Haplotypes nested in 1-4	1-19-20-2-11-12-NO	Contiguous range expansion
Haplotypes nested in 1-6	1-12-11-12-NO	Contiguous range expansion
Haplotypes nested in 1-7	1-2-3-4-9-NO	Allopatric fragmentation
One-step clades nested in 2-2	1-2-11-12-NO	Contiguous range expansion

[a] Haplotypes nested in clade 1-2 and higher clades 2-1, 2-5, 3-1, and 3-2 are not presented in this table because they yield a nonconclusive inference.
[b] Inference chains are obtained using the inference key (31, 43) throughout the hierarchical clade levels.
[c] Evolution processes are deduced from the inference chain (31, 43).

PPV does not easily invade "occupied land," making conciliation with our results possible. Indeed, under our experimental conditions, with the virus being present well before expansion of new leaves, we propose that the cells initiating buds might become rapidly infected by very few virus genomes (most often a single one) present in neighboring cells, subsequently precluding secondary infection by other variants. In this hypothesis, newly formed tissues are colonized by locally established haplotypes and not by those circulating in the phloem, as one would expect from the general scheme that describes virus primary invasion of a healthy host (38).

H1 was the only haplotype detected in the bark of the trunk. Interestingly, the bark samples of the major limbs contain H1 and three additional haplotypes, the bark samples of yearling branches contain the four haplotypes detected below plus two additional ones, and the leaves contain the 6 detected below plus 27 new ones. This addition of new haplotypes when going from old to young organs suggests that the tree reflects the chronology of the appearance of virus diversity. Consistently, the phylogeny network containing all haplotypes (Fig. 3) proposes that H1 is the most probable root from which diversity has built up slowly over the years. Although indirect, these data partially compensate for the fact that the original virus source plant, on which the aphids were fed prior to inoculation of the tree analyzed here, was lost 13 years ago, preventing any possibility of directly identifying the ancestral haplotype. H1 was also found in old main roots, whereas additional haplotypes appeared in young terminal roots, perhaps suggesting a similar phenomenon of PPV diversification downward with the growth of the root system. The presence (together with H1) of some haplotypes at a low frequency (below 0.1) in the initial inoculum, and their ulterior increase in frequency due to genetic drift and/or founder effect in samples of limbs, branches, and leaves, cannot be excluded. However, because the virus has replicated for 13 years in the tree and because the haplotypic diversity increases in comparable younger tissues (see H and π in barks of trunk, barks of limbs, and barks of branches in Table 2), it is likely that most haplotypes were derived from H1, accumulating new mutations after the original inoculation event.

As mentioned in above, fragmentary data had previously hinted at a within-host differentiation of virus populations in perennial plants. AMOVA based on the breakdown of our data set into 30 discrete subpopulations confirms the importance of this phenomenon for PPV in a *Prunus* host. More innovatively, AMOVA demonstrates that this differentiation is structured by the architecture of the tree, with subpopulations diverging most significantly in different major limbs. From Fig. 2, one can immediately perceive that different PPV lines become isolated in majors limbs, with haplotypes that occur frequently in samples from one limb, with its associated branches and leaves, being totally absent in samples from another limb. To statistically sustain this hypothesis, we applied NCA to evaluate the association between haplotype and geographical range within the host. First of all, the phylogeny network shows that haplotypes abundant in more than one major limb (H1 and H4) are likely the most ancestral, suggesting that they existed prior to limb development and corresponding PPV line isolation. Moreover, the inferred evolution process for clades 1-1, 1-4, 1-6, and 2-2 is contiguous range expansion, defined previously by Templeton (41) as a gradual, moving front of expansion for a given population. The possibility of range expansion for a virus population in a tree that is already systemically infected is limited to the newly formed branches and leaves (and roots) appearing over time. This result is consistent with the observation that PPV haplotypes seem to be unable to recolonize previously infected cells and confirms our conclusion that the virus population is expanding in one direction, toward young extremities as they develop, with no possibility of return and mixing via the vascular system in previously infected organs and tissues. We believe that this mechanism satisfactorily explains how several lines become isolated in different parts of the tree. Because neutrality tests detected no selection acting on any haplotypes, it is also likely that upon isolation, the lines differentiate rapidly due to the drastic bottlenecks discussed above and the associated genetic drift. Contiguous range expansion is the most likely inference, as it concerned 282 of the 333 samples analyzed. Other clades gave inconclusive inferences, except clade 1-7, the evolutionary process of which was inferred as allopatric fragmentation. We believe that this latter case is not reliable since clade 1-7 groups a very limited number of samples.

Overall, the results presented in this report have important implications, as they challenge some views commonly adopted among plant virologists and beyond. While panmixy is being denied for within-host HIV populations, with the concept of the metapopulation applying better, we demonstrate here that differentiation can go further in other virus species such as PPV, with several populations becoming isolated and evolving independently for years in a single host. PPV is a member of the genus *Potyvirus*, which alone accounts for nearly 25% of described plant virus species. It will be interesting to evaluate how the situation reported here applies to other potyviruses and to other virus genera infecting annual or perennial plants. Finally, the fact that the leaves harbor a tremendous collection of clones from the virus population(s) is also a striking situation that has been largely overlooked, with huge implications for both theoretical and practical aspects of virus biology, epidemiology, and evolution; how these findings apply to other tree-infecting virus species appears an interesting prospect, as it is, in large majority, from the leaves that viruses are acquired during transmission by aerial vectors.

ACKNOWLEDGMENTS

We thank Fabien Halkett and the group CaGeTE for critical reading of the manuscript.

This work was supported by the government of Tunisia, by the French ministry of Agriculture, and by a grant from DADP-INRA/Région Languedoc-Roussillon.

REFERENCES

1. **Cabot, B., M. Martell, J. I. Esteban, S. Sauleda, T. Otero, R. Esteban, J. Guardia, and J. Gomez.** 2000. Nucleotide and amino acid complexity of hepatitis C virus quasispecies in serum and liver. J. Virol. **74:**805–811.
2. **Clement, M., D. Posada, and K. A. Crandall.** 2000. TCS: a computer program to estimate gene genealogies. Mol. Ecol. **9:**1657–1659.
3. **Crandall, K. A.** 1996. Multiple interspecies transmissions of human and simian T-cell leukemia/lymphoma virus type I sequences. Mol. Biol. Evol. **13:**115–131.
4. **Crandall, K. A., and A. R. Templeton.** 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics **134:**959–969.
5. **Dietrich, C., and E. Maiss.** 2003. Fluorescent labelling reveals spatial separation of potyvirus populations in mixed infected Nicotiana benthamiana plants. J. Gen. Virol. **84:**2871–2876.

6. **Domingo, E.** 2002. Quasispecies theory in virology. J. Virol. **76:**463–465.
7. **Domingo, E., and J. J. Holland.** 1997. RNA virus mutations and fitness for survival. Annu. Rev. Microbiol. **51:**151–178.
8. **Drake, J. W., and J. J. Holland.** 1999. Mutation rates among RNA viruses. Proc. Natl. Acad. Sci. USA **96:**13910–13913.
9. **Ducoulombier, D., A. M. Roque-Afonso, G. Di Liberto, F. Penin, R. Kara, Y. Richard, E. Dussaix, and C. Feray.** 2004. Frequent compartmentalization of hepatitis C virus variants in circulating B cells and monocytes. Hepatology **39:**817–825.
10. **D'Urso, F., M. A. Ayllòn, L. Rubio, A. Sambade, A. Hermeso de Mendoza, J. Guerri, and P. Moreno.** 2000. Contribution of uneven distribution of genomic RNA variants of Citrus tristeza virus (CTV) within the plant to changes in the viral population following aphid transmission. Plant Pathol. **49:**288–294.
11. **D'Urso, F., A. Sambade, A. Moya, J. Guerri, and P. Moreno.** 2003. Variation of haplotype distributions of two genomic regions of Citrus tristeza virus populations from eastern Spain. Mol. Ecol. **12:**517–526.
12. **Eigen, M.** 1996. On the nature of virus quasispecies. Trends Microbiol. **4:**216–218.
13. **Excoffier, L., P. E. Smouse, and J. M. Quattro.** 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics **131:**479–491.
14. **Fargette, D., A. Pinel, Z. Abubakar, O. Traore, C. Brugidou, S. Fatogoma, E. Hebrard, M. Choisy, Y. Sere, C. Fauquet, and G. Konate.** 2004. Inferring the evolutionary history of *Rice yellow mottle virus* from genomic, phylogenetic, and phylogeographic studies. J. Virol. **78:**3252–3261.
15. **French, R., and D. C. Stenger.** 2003. Evolution of Wheat streak mosaic virus: dynamics of population growth within plants may explain limited variation. Annu. Rev. Phytopathol. **41:**199–214.
16. **Froissart, R., C. O. Wilke, R. Montville, S. K. Remold, L. Chao, and P. E. Turner.** 2004. Co-infection weakens selection against epistatic mutations in RNA viruses. Genetics **168:**9–19.
17. **Frost, S. D., M. J. Dumaurier, S. Wain-Hobson, and A. J. Brown.** 2001. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. Proc. Natl. Acad. Sci. USA **98:**6975–6980.
18. **Garcia-Arenal, F., A. Fraile, and J. M. Malpica.** 2001. Variability and genetic structure of plant virus population. Annu. Rev. Phytopathol. **39:**157–186.
19. **Glasa, M., V. Marie-Jeanne, G. Labonne, Z. Subr, O. Kudela, and J. B. Quiot.** 2002. A natural population of recombinant Plum pox virus is viable and competitive under field conditions. Eur. J. Plant Pathol. **108:**843–853.
20. **Hall, J. S., R. French, G. L. Hein, T. J. Morris, and D. C. Stenger.** 2001. Three distinct mechanisms facilitate genetic isolation of sympatric wheat streak mosaic virus lineages. Virology **282:**230–236.
21. **Holmes, E. C.** 2003. Error thresholds and the constraints to RNA virus evolution. Trends Microbiol. **11:**543–546.
22. **Holmes, E. C., and A. Moya.** 2002. Is the quasispecies concept relevant to RNA viruses? J. Virol. **76:**460–462.
23. **Hudson, R. R.** 1989. How often are polymorphic restriction sites due to a single mutation? Theor. Popul. Biol. **36:**23–33.
24. **Kemal, K. S., B. Foley, H. Burger, K. Anastos, H. Minkoff, C. Kitchen, S. M. Philpott, W. Gao, E. Robison, S. Holman, C. Dehner, S. Beck, W. A. Meyer III, A. Landay, A. Kovacs, J. Bremer, and B. Weiser.** 2003. HIV-1 in genital tract and plasma of women: compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc. Natl. Acad. Sci. USA **100:**12972–12977.
25. **Laskus, T., M. Radkowski, L. F. Wang, M. Nowicki, and J. Rakela.** 2000. Uneven distribution of hepatitis C virus quasispecies in tissues from subjects with end-stage liver disease: confounding effect of viral adsorption and mounting evidence for the presence of low-level extrahepatic replication. J. Virol. **74:**1014–1017.
26. **Li, H., and M. J. Roossinck.** 2004. Genetic bottlenecks reduce population variation in an experimental RNA virus population. J. Virol. **78:**10582–10587.
27. **Magome, H., N. Yoshikawa, and T. Takahashi.** 1998. Single-strand conformation polymorphism analysis of Apple stem grooving capillovirus sequence variants. Phytopathology **89:**136–140.
28. **Moya, A., S. F. Elena, A. Bracho, R. Miralles, and E. Barrio.** 2000. The evolution of RNA viruses: a population genetics view. Proc. Natl. Acad. Sci. USA **97:**6967–6973.
29. **Okamoto, H., T. Nishizawa, M. Takahashi, S. Asabe, F. Tsuda, and A. Yoshikawa.** 2001. Heterogeneous distribution of TT virus of distinct genotypes in multiple tissues from infected humans. Virology **288:**358–368.
30. **Philpott, S., H. Burger, C. Tsoukas, B. Foley, K. Anastos, C. Kitchen, and B. Weiser.** 2005. Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and intrapatient recombination. J. Virol. **79:**353–363.
31. **Posada, D., and T. R. Buckley.** 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. **53:**793–808.
32. **Posada, D., K. A. Crandall, and A. R. Templeton.** 2000. GeoDis: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. Mol. Ecol. **9:**487–488.
33. **Potter, S. J., P. Lemey, G. Achaz, C. B. Chew, A. M. Vandamme, D. E. Dwyer, and N. K. Saksena.** 2004. HIV-1 compartmentalization in diverse leukocyte populations during antiretroviral therapy. J. Leukoc. Biol. **76:**562–570.
34. **Rozas, J., and R. Rozas.** 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics **15:**174–175.
35. **Rubio, L., J. Guerri, and P. Moreno.** 2000. Presented at the 14th Conference of the International Organization of Citrus Virologists, Riverside, Calif.
36. **Sacristan, S., J. M. Malpica, A. Fraile, and F. Garcia-Arenal.** 2003. Estimation of population bottlenecks during movement of *Tobacco mosaic virus* in tobacco plants. J. Virol. **77:**9906–9911.
37. **Sanjuan, R., F. M. Codoner, A. Moya, and S. F. Elena.** 2004. Natural selection and the organ-specific differentiation of HIV-1 V3 hypervariable region. Evol. Int. J. Org. Evol. **58:**1185–1194.
38. **Santa Cruz, S.** 1999. Perspective: phloem transport of viruses and macromolecules—what goes in must come out. Trends Microbiol. **7:**237–241.
39. **Simmonds, P.** 2004. RNA viruses—evolution in action. Microbiol. Today **31:**163–165.
40. **Templeton, A. R.** 1992. Human origins and analysis of mitochondrial DNA sequences. Science **255:**737.
41. **Templeton, A. R.** 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Mol. Ecol. **7:**381–397.
42. **Templeton, A. R.** 1987. Nonparametric phylogenetic inference from restriction cleavage sites. Mol. Biol. Evol. **4:**315–323.
43. **Templeton, A. R., E. Routman, and C. A. Phillips.** 1995. Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genetics **140:**767–782.
44. **Templeton, A. R., C. F. Sing, A. Kessling, and S. Humphries.** 1988. A cladistic analysis of phenotype associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. Genetics **120:**1145–1154.
45. **Vigne, E., M. Bergdoll, S. Guyader, and M. Fuchs.** 2004. Population structure and genetic variability within isolates of Grapevine fanleaf virus from a naturally infected vineyard in France: evidence for mixed infection and recombination. J Gen. Virol. **85:**2435–2445.
46. **Weir, B. S., and C. C. Cockerham.** 1984. Estimating F-statistics for the analysis of population structure. Evolution **38:**1358–1370.
47. **Wilke, C. O.** 2005. Quasispecies theory in the context of population genetics. BMC Evol. Biol. **5:**44.